

Les tableaux de fréquences 2×2 et leur traitement statistique

L. Laurencelle

Université du Québec à Trois-Rivières

Fait peu connu des chercheurs, voire des statisticiens appliqués, les tableaux de fréquences 2×2 appartiennent à trois catégories logiques différentes, selon que les effectifs totaux des rangées et des colonnes soient fixés d'avance ou non. Dans le modèle sans prédétermination, ou modèle 0, chaque élément peut s'inscrire dans n'importe quelle rangée ou colonne, comme lorsqu'on croise deux items dichotomiques d'un questionnaire. N'ayant trouvé aucun traitement spécifique de ce modèle dans la documentation, nous en proposons un, l'indice de concordance C , vérifié statistiquement par un simple test binomial. La comparaison des proportions obtenues dans deux groupes représente le modèle 1, les tailles de groupes étant prédéterminées : un test dû à Liddell et son approximation par le Khi-deux sont proposés, justifiés et validés. Enfin, dans le modèle 2, le total des réponses d'une sorte et de l'autre est connu d'avance ; pour ces cas, plus rares, le « test exact de Fisher » sied parfaitement, et le Khi-deux de Yates (avec correction de $-\frac{1}{2}n$) en est l'approximation par excellence. Les tests basés sur les rapports de vraisemblance ainsi que d'autres approximations sont aussi discutés.

Little known to researchers as well as applied statisticians, 2×2 frequency tables can be classified into three logical categories, depending on whether row and column totals are fixed or not. In model 0, with nothing predetermined (except total sample size n), each element may be tallied up in any row or column, as when we correlate two dichotomous questions in a survey. No specific treatment of model 0 being found in the literature, we propose one, the index of agreement C , assessed quite simply through a binomial test. Comparing two proportions issuing from two groups comes under model 1, group sizes being predetermined: a procedure due to Liddell and its Chi-square approximation are laid down, justified and validated. Finally, in model 2, response totals are known in advance in both dimensions. For these rarer cases, « Fisher's exact test » suits perfectly, with its quasi perfect approximation by Yates-modified Chi-square formula (with the $-\frac{1}{2}n$ correction). Tests based on likelihood ratios and other approximate procedures are also reviewed.

Il existe une pléthore de procédures de test statistique qui ont été proposées et qui sont disponibles pour le traitement d'un tableau de fréquences 2×2 alors que, à première vue, tous les tableaux 2×2 se ressemblent désespérément l'un l'autre. Y a-t-il des différences essentielles, des catégories de tableaux 2×2 ? Quels sont les principaux tests statistiques

disponibles? Parmi ces tests, lequel ou lesquels sont appropriés pour chaque catégorie de tableaux? Le but de cette note est donc de poser clairement le problème du traitement des tableaux 2×2 et d'en examiner méthodiquement les solutions.

Les tableaux 2×2 et leurs catégories

Le tableau de fréquences 2×2 dénote la répartition de n éléments, individus ou événements dans un tableau à deux dimensions binaires. Deux notations sont couramment

L. Laurencelle, Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières, Québec, G9A 5H7, CANADA.

	Y+	Y-		Groupe 1	Groupe 2	
X+	a	b	+	n ₁₁	n ₁₂	n _{1.}
X-	c	d	-	n ₂₁	n ₂₂	n _{2.}
	n			n _{1.}	n _{2.}	n

Figure 1. Dispositions et notations possibles d'un tableau de fréquences 2 x 2

utilisées (voir figure 1).

Le tableau présente généralement n personnes (ou objets) réparties dans les quatre cellules, selon deux dimensions : les rangées et les colonnes. D'un point de vue statistique, c.-à-d. en considérant l'espace de variation (ou les degrés de liberté) qui caractérise chaque tableau, on distingue trois catégories, disons trois modèles (Cochran 1952 ; Kendall et Stuart 1979 ; Sokal et Rohlf 1981)¹.

Le modèle 0 : Les n sujets se répartissent *ad lib* selon les rangées ou les colonnes, aucun total n'étant prédéterminé. Le tableau comporte 3 degrés de liberté, soit les 4 effectifs dans les cellules, moins l'effectif total n qui est prédéterminé.

Le modèle 1 : Les totaux d'une dimension, p.ex. les colonnes, sont prédéterminés et constituent des groupes, les sujets de chaque groupe pouvant opter selon l'autre dimension, p.ex. les rangées. Le tableau comporte donc 2 degrés de liberté.

Le modèle 2 : Les totaux des deux dimensions sont prédéterminés, les sujets ne pouvant opter que dans l'observance de ces totaux. Le tableau comporte seulement 1 degré de liberté.

Exemples

Modèle 0. La situation type qui génère un tableau de modèle 0, sans totaux marginaux prédéterminés, se retrouve dans les sondages et l'analyse des questionnaires : il s'agit ici d'étudier comment la répartition des sujets dans les catégories d'une dimension (X) interagit avec celle dans l'autre dimension (Y), tel qu'illustré à la figure 1, à gauche.

Supposons une enquête auprès de $n = 50$ participants ; deux questions posées portent sur la raison première des choix alimentaires (Plaisir *vs* Santé) et sur le type préféré de loisirs (Culture *vs* Sport), avec la répartition présentée à la figure 2 (données fictives) :

	Culture	Sport
Plaisir	28	2
Santé	12	8

Figure 2. Exemple numérique du modèle 0

A priori, les $n = 50$ répondants avaient l'option de se loger dans n'importe laquelle des 4 cellules. Le nombre d'arrangements possibles correspond ainsi au nombre de configurations différentes de n objets en 4 cellules. Disposant les n objets en une rangée, et plaçant les 3 bornes inter-cellules à différents endroits de la rangée, nous voyons qu'il en résulte ${}_{n+3}C_3 = (n+1)(n+2)(n+3)/6$ arrangements possibles² (23 426 dans le cas présent), soit une combinatoire d'ordre n^3 .

La question posée ici est de savoir s'il y a interaction, ou corrélation (au sens large), entre les deux réponses.

Modèle 1. Le modèle 1, illustré par le tableau de droite de la figure 1, représente sans doute l'occurrence la plus répandue et la mieux documentée des tableaux 2 x 2. Ce modèle équivaut, à toutes fins pratiques, à la comparaison de deux proportions indépendantes, selon la différence :

$$p_1 - p_2 = n_{11} / n_{1.} - n_{12} / n_{2.}, \quad (1)$$

différence qu'on pourrait aussi écrire « $a / (a+c) - b / (b+d)$ ». Prenons l'exemple d'une petite enquête menée auprès de 10 Hommes et 10 Femmes, qui devaient exprimer leur préférence quant aux couleurs (mates ou brillantes) de leur environnement de travail, tel qu'on le voit à la figure 3.

	Hommes	Femmes
Couleurs mates	6	2
Couleurs brillantes	4	8

Figure 3. Exemple numérique du modèle 1

Les n_1 sujets du groupe 1 ont théoriquement la possibilité d'opter pour une rangée (+) ou l'autre (-), les répartitions possibles étant donc $(0, n_1), (1, n_1 - 1), \dots, (n_1, 0)$, au nombre total de $(n_1 + 1)$; de même pour les n_2 sujets du groupe 2. Le tableau de modèle 1 offre donc $(n_1 + 1)(n_2 + 1)$ configurations aléatoires, soit une combinatoire d'ordre n^2 , égale ici à 121.

² De ces arrangements, il faut cependant retrancher ceux pour lesquels l'une des dimensions est dégénérée, c.-à-d. réduite à une seule valeur. Or, il existe $4n$ arrangements pour lesquels au moins l'une des sommes $(a+b), (c+d), (a+c)$ ou $(b+d)$ est nulle, d'où la combinatoire réelle de ce modèle est de ${}_{n+3}C_3 - 4n$.

¹ J'ai numéroté les modèles par les chiffres 0, 1 et 2 afin que le chiffre corresponde au nombre de dimensions prédéterminées (ou totaux marginaux fixés d'avance) du tableau. La documentation étiquette ces modèles de différentes manières.

Existe-t-il une différence sérieuse de préférence entre les sexes, ou la différence observée de 0,4 (= 0,6 - 0,2) est-elle réductible au hasard?

Modèle 2. Dans les tableaux du modèle 2, d'occurrence plus rare, la valeur des totaux marginaux est fixée d'avance. L'exemple princeps d'un tel tableau est le suivant³. On présente à un sujet 40 petits biscuits, dont 15 sont tartinés de beurre et 25 de margarine : le sujet doit identifier l'ingrédient, *ce en respectant les nombres indiqués* (15 et 25) pour chacun. Le résultat donné apparaît à la figure 4.

	Beurre	Margarine
Goût de beurre	4	11
Goût de margarine	11	14

Figure 4. Exemple numérique du modèle 2

Un autre type de tableau de ce modèle serait obtenu par l'élection de n_1 représentants dans un groupe constitué de sous-groupes de n_{11} et n_{12} personnes : p. ex., un syndicat formé de 8 hommes et 20 femmes désigne un exécutif où se retrouvent 3 hommes et 1 femme.

Les totaux des rangées et colonnes étant fixés, il reste un seul degré de liberté dans le tableau, une seule cellule libre. En fait, soit $t = \min(a+b, c+d, a+c, b+d)$, ici $t = 15$: la plus petite cellule a pour champ de variation 0, 1, ..., $t < n$, d'où, grosso modo, une combinatoire d'ordre n .

Les types de tartines perçus correspondent-ils à la réalité?

Et, finalement, est-il vraisemblable qu'une seule procédure de test convienne à tous les tableaux 2×2 présentés, quel qu'en soit le modèle?

Survole des tests statistiques disponibles

Les manuels de statistique appliquée proposent une panoplie de procédures pour arriver à une décision statistique à partir d'un tableau de fréquences 2×2 , et la plupart le font sans distinction pour le modèle de variation qu'il met en œuvre. D'une part, cette négligence du modèle trouve une excuse et une explication dans l'affirmation de Kendall et Stuart (1979, p. 584-585) à l'effet que le traitement du modèle 2 (par le test de Fisher, voir plus bas) est aussi « optimum » pour les deux autres modèles de tableau. D'autre part, sauf peut-être à titre d'approximation ou au niveau asymptotique, la procédure qui serait adéquate pour un modèle peut difficilement le rester pour un autre.

La procédure de test la plus connue est celle du Khi-deux

simple⁴ (Howell 1998), distribué selon la loi χ^2 à 1 degré de liberté soit, à partir de la forme (a, b, c, d) de la figure 1,

$$\text{Khi-deux brut} = \frac{n(a \cdot d - b \cdot c)^2}{(a+b)(c+d)(a+c)(b+d)} \quad (2)$$

Apparaît aussi fréquemment le Khi-deux avec correction de Yates (Siegel et Castellan 1988), soit :

$$\text{Khi-deux (Yates)} = \frac{n(|a \cdot d - b \cdot c| - \frac{1}{2}n)^2}{(a+b)(c+d)(a+c)(b+d)} \quad (3)$$

Certains auteurs introduisent le test de Fisher (Siegel et Castellan 1988), parfois sous le nom « test des probabilités exactes de Fisher », à titre d'étalon numérique pour les autres tests. Soit le tableau (a, b, c, d) et la probabilité hypergéométrique

$$p_{\text{hyp}}(a, b, c, d) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n! a! b! c! d!} \quad (4)$$

alors la probabilité associée au test de Fisher est :

$$p(\text{Fisher}) = \sum_{(a', b', c', d')} p_{\text{hyp}}(a', b', c', d') \quad (5)$$

la sommation se faisant sur les configurations (a', b', c', d') plus extrêmes par rapport à celle observée.

Le test G (Sokal et Rohlf 1981), ou test du rapport de vraisemblance, imite en quelque sorte le test de Fisher, à partir d'une approximation du logarithme de la vraisemblance par la loi du χ^2 . La formule suivante encapsule les calculs requis :

$$G = 2 \{ a \ln a + b \ln b + c \ln c + d \ln d + n \ln n - (a+b) \ln(a+b) - (c+d) \ln(c+d) - (a+c) \ln(a+c) - (b+d) \ln(b+d) \} \quad (6)$$

la statistique ayant pour distribution approximative la loi χ^2 avec 1 degré de liberté.

On trouve d'autres formes de statistiques applicables, notamment un test z sur la différence entre deux proportions, et des variantes des formes mentionnées ci-dessus. Nous les présenterons au moment opportun.

Traitement statistique du modèle 2 : le test de Fisher et son approximation

Théorie du traitement. Dans le modèle 2, les totaux des rangées et colonnes sont prédéterminés et la variation des effectifs du tableau (a, b, c, d) doit obéir aux contraintes d'un échantillonnage sans remise. Selon l'hypothèse nulle (de non différence), la probabilité qu'un élément se retrouve en X+

³ Repris de Kendall et Stuart (1979), p. 581, qui le tire lui-même de R. A. Fisher.

⁴ La formule de calcul présentée équivaut à la forme générale du Khi-deux, soit $\sum (f_{i,j} - ft_{i,j})^2 / ft_{i,j}$, consistant à comparer les fréquences (ou effectifs) obtenus $f_{i,j}$ aux fréquences théoriques $ft_{i,j}$, i et j désignant respectivement la rangée et la colonne du tableau.

(plutôt qu'en X^-) n'est pas constante, car elle dépend du nombre d'éléments encore disponibles et elle varie avec eux : au lieu d'une probabilité binomiale, nous avons affaire ici à une probabilité hypergéométrique.

En fait, la probabilité de chaque arrangement (a, b, c, d) doit être considérée selon le nombre d'arrangements conjoints possibles en X et en Y , soit :

$$p(a, b, c, d) = \frac{\binom{n}{a,b,c,d}}{\binom{n}{a+b} \binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

tel qu'indiqué en (4). La probabilité obtenue caractérise le tableau (a,b,c,d) de départ. Pour effectuer le test, les probabilités $p(a',b',c',d')$ attachées aux variations plus extrêmes du tableau doivent être ajoutées, produisant la probabilité-test, p_{hyp} , avec un résultat significatif si cette dernière n'excède pas le seuil de probabilité fixé, soit $p_{hyp} \leq \alpha^5$.

Un exemple numérique. Appliquons la procédure décrite aux données de beurre et margarine de la figure 4, avec le tableau (4, 11, 11, 14). D'abord, nous obtenons $p(4,11,11,14) = 15! 25! 15! 25! / [40! 4! 11! 11! 14!] = 0,1513$, la margarine étant mieux identifiée ($14 / 25 = 0,56$) que le beurre ($4 / 15 \approx 0,27$). Les variations plus extrêmes correspondent aux tableaux $(a',b',c',d') = (3,12,12,13), (2,13,13,12), (1,14,14,11)$ et $(0,15,15,10)$, de probabilités respectives 0,0588, 0,0136, 0,0017 et 0,0001, pour une somme globale de $p_{hyp} \approx 0,2255$.

Approximation. C'est pour approximer cette probabilité par la loi du χ^2 que Yates (1934) a proposé sa fameuse « correction de continuité », de même inspiration que celle utilisée pour l'approximation d'une probabilité binomiale par la loi normale (voir (18), plus loin). Ainsi, appliquant la formule (3) du Khi-deux, nous obtenons :

$$\text{Khi-deux(Yates)} = \frac{40 \times (|4 \times 14 - 11 \times 11| - 40 / 2)^2}{15 \times 25 \times 15 \times 25} = 0,576,$$

valeur dont la probabilité extrême (à droite) est d'environ 0,221. Plusieurs auteurs (Conover 1974, Liddell 1978, Pirie et Hamdan 1972) concordent pour affirmer la valeur du Khi-deux avec correction de « $-n/2$ » comme meilleure approximation du test de Fisher. Howell (1998) lui-même, qui ne recommande pourtant pas le Khi-deux de Yates (parce qu'il prétend que son application est trop rare ou mal comprise), le confirme comme étant approprié pour le

⁵ Tout au long de cet article, pour simplifier l'exposé et sans perte de généralité, nous nous confinerons à des hypothèses et des tests *unilatéraux*.

traitement des tableaux de modèle 2. Haber (1980) arrive à d'autres conclusions en comparant les Khi-deux brut et corrigé à trois autres approximations, chacune assez laborieuse, ces nouvelles procédures lui semblant préférables. Nous revenons sur ce sujet plus loin.

Étude des approximations. Malgré que le test de Fisher constitue la solution correcte pour l'analyse d'un tableau répondant au modèle 2, la littérature a développé et entretenu, parfois dans une certaine confusion, d'autres procédures permettant de sanctionner la significativité des données. Il faut dire que les calculs nécessaires au test de Fisher deviennent prohibitifs lorsque la taille (*p. ex.* $a+b+c+d$) augmente, motivant ainsi le recours à une procédure plus simple. Nous avons donc retenu quatre approximations plus ou moins classiques du test de Fisher, soit :

- le Khi-deux brut (formule 2), qui vaut 1,202 pour notre exemple ;
- le Khi-deux de Yates (formule 3), valant ici 0,576 ;
- le G (ou test du logarithme de la vraisemblance) (formule 6), ici égal à 1,231 ;
- le G_{corr} , une version du G améliorée par la correction de Williams (1976) définie ci-dessous (voir aussi Sokal et Rohlf 1981) :

$$G_{corr} = G / q \quad (7)$$

$$\text{où } q = 1 + \frac{\left(\frac{n}{a+b} + \frac{n}{c+d} - 1\right) \left(\frac{n}{a+c} + \frac{n}{b+d} - 1\right)}{6n}$$

Sur les données de la figure 4, $q \approx 1,0445$ et $G_{corr} \approx 1,179$.

Pour effectuer la comparaison, nous avons exploré tous les tableaux générés à partir de groupes de tailles n_1 et n_2 , c.-à-d. $(n_1+1)(n_2+1) - 2$ tableaux, ce pour différentes valeurs de n_1 et n_2 . Les figures 5 gauche et 5 droite font voir les différences absolues moyennes entre la probabilité (unilatérale) calculée pour le test de Fisher et celle obtenue pour chacune des quatre approximations.

Quel que soit son mérite propre, on voit, à la figure 5 gauche, que le Khi-deux de Yates se démarque avantageusement des autres approximations, ces dernières montrant un comportement mutuellement semblable. Si, comme à la figure 5 droite, on restreint le calcul aux seuls cas où au moins l'une des probabilités comparées atteint le seuil de 0,05, alors le Khi-deux de Yates, avec correction de continuité, apparaît quasi parfait dès les tailles modestes de $n_1 = n_2 = 5$ ou 10, les approximations finissant par le rejoindre pour des tailles plus élevées. Quant aux tableaux à tailles inégales, l'exemple extrême comportant $n_1 = 5$ et $n_2 = 55$ éléments peut suffire : on y obtient moy $|p_{\text{Khi2(Yates)}} - p_{\text{Fisher}}| \approx 0,0033$ pour le Khi-deux de Yates, les autres statistiques étant de 0,0229 (Khi2), 0,0308 (G) et 0,0212 (G_{corr}).

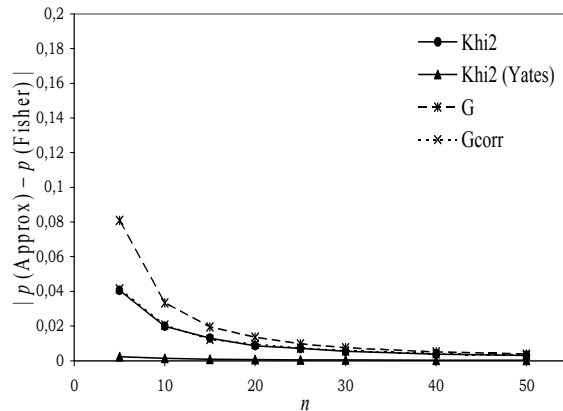
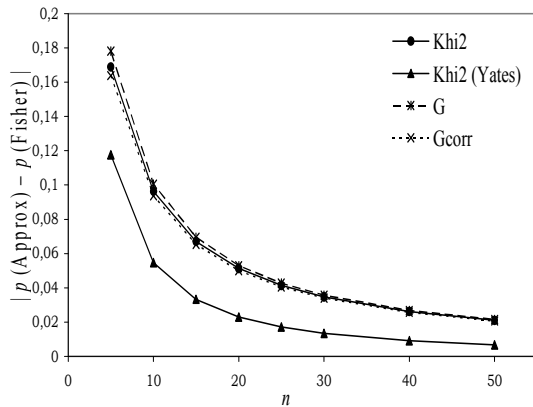


Figure 5 (gauche). Erreur moyenne en probabilité de 4 approximateurs du test de Fisher ($n_1 = n_2 = n$, tous p).

Figure 5 (droite). Erreur moyenne en probabilité de 4 approximateurs du test de Fisher ($n_1 = n_2 = n$, $p < 0,05$)

Les figures 6 gauche et 6 droite font voir un autre aspect de la qualité des approximations du test de Fisher. Dans ces figures, seuls sont retenus les cas où soit le test de Fisher, soit le test approximatif atteint la significativité (au seuil unilatéral de 0,05), et le nombre de cas de discordances est tracé : les figures illustrent donc la proportion de discordances⁶ de chaque approximateur, selon n .

La qualité exceptionnelle d'approximation du Khi-deux de Yates ressort clairement des tracés de la figure 6 gauche, les autres procédures s'améliorant lentement selon n croissant. À la figure 6 droite, où sont comparés des groupes à différence de tailles croissante, le Khi-deux de Yates maintient sa valeur, même pour des groupes à $n_1 = 5$ et $n_2 = 55$, alors que la qualité des autres approximations est au mieux médiocre.

Haber (1980), on l'a dit, propose des procédures, inspirées de Cochran et d'autres, dont le but était d'améliorer l'approximation du test de Fisher au-delà de celle obtenue par le Khi-deux de Yates. Nous avons tenté de valider ses propositions en examinant tous les cas, peu nombreux, de discordances apparaissant aux figures 6 gauche et 6 droite, soit un total de 34 cas. Il ressort de l'examen que, dans tous les cas sauf ceux qu'elles laissent concordants par rapport au Khi-deux de Yates, les procédures⁷ proposées par Haber dégradent en fait l'approximation⁸.

⁶ Par discordances, sont compris tous les cas où le test de Fisher et le test approximatif se contredisent, l'un étant significatif (avec $p < 0,05$) et l'autre non. La proportion affichée est obtenue en divisant les discordances recensées par le nombre de configurations générées, soit $(n_1+1)(n_2+1) - 2$.

⁷ Il s'agit des méthodes dites C et S, inspirées de W.G. Cochran, et de la méthode M tirée de N. Mantel, voir Haber (1980), p. 511.

⁸ Haber (1980) effectue son évaluation des procédures dans le contexte de tests bilatéraux. Il est possible, ou c'est du moins notre impression, que certaines procédures « améliorées »

Pour conclure sur le traitement des tableaux à totaux prédéterminés des rangées et des colonnes, le test de Fisher est bel et bien la procédure appropriée, et le Khi-deux amendé par Yates (1934), avec sa correction de « $-n/2$ », en fournit la meilleure approximation.

Traitement statistique du modèle 1 : la différence entre deux groupes indépendants

Théorie du traitement. Les effectifs d'une seule dimension sont connus à l'avance dans les situations correspondant au modèle 1 : à toutes fins pratiques, le traitement du tableau revient à la comparaison des réponses dichotomiques dans deux groupes, c.-à-d. au test de la différence entre les deux proportions p_1 et p_2 , où $p_j = n_{1j} / n_{.j}$, basées respectivement sur les groupes de $n_{.1}$ et $n_{.2}$ éléments. Sous l'hypothèse d'indépendance statistique des groupes, chaque élément d'un groupe j va opter pour la catégorie cible (+) avec une probabilité π_j , engendrant une répartition (n_{1j}, n_{2j}) de probabilités binomiales $p(n_{1j} | \pi_j, n_{.j})$. De plus, sous l'hypothèse de non différence, les probabilités π_1 et π_2 sont égales, soit $\pi_1 = \pi_2 = \pi_0$. Connaissant π_0 , la probabilité associée à la différence

$$p_1 - p_2 = n_{11} / n_{.1} - n_{12} / n_{.2}$$

est simplement le produit des probabilités respectives de ses ingrédients, soit :

$$p(p_1 - p_2 | \pi_0) = p(n_{11} | \pi_0, n_{.1}) \times p(n_{12} | \pi_0, n_{.2})$$

utilisent une compensation controlatérale, notamment pour le test de Fisher lui-même, où la probabilité calculée est celle (souvent asymétrique) attachée à toute différence égale ou plus grande par rapport à celle observée, quel que soit le signe de cette différence. Notre contexte de test unilatéral n'autorise pas un tel calcul, calcul dont nous suspectons en tout état de cause la légitimité.

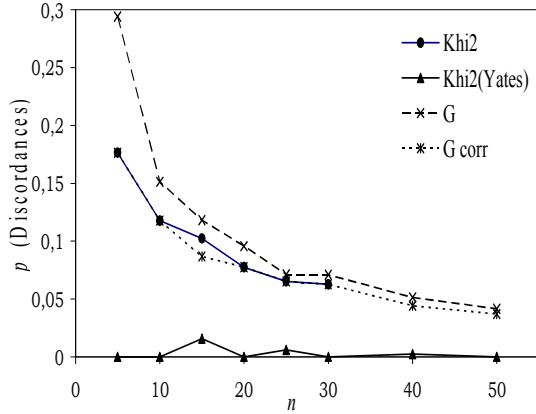


Figure 6 (gauche). Discordances de 4 approximeurs du test de Fisher ($n_1 = n_2 = n$).

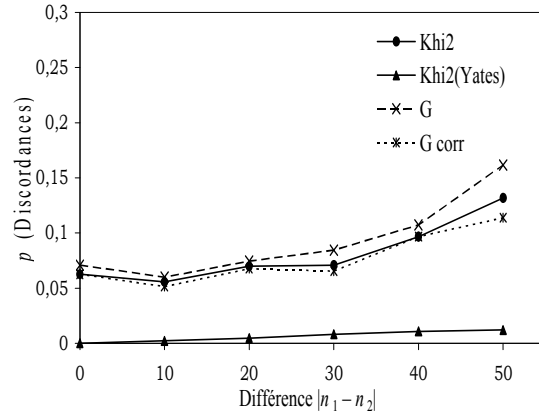


Figure 6 (droite). Discordances de 4 approximeurs du test de Fisher ($n_1 \neq n_2, n_1 + n_2 = 60$)

et la probabilité-test s'obtient en additionnant les probabilités associées aux résultats binomiaux égaux ou plus extrêmes, soit à l'ensemble A tel que :

$$A = \{ n_{11}', n_{12}', n_{11}'/n_{.1} - n_{12}'/n_{.2} \geq n_{11}/n_{.1} - n_{12}/n_{.2} \}.$$

Ce calcul, qu'on retrace notamment dans Kendall et Stuart (1979), suppose connue la valeur du paramètre commun de probabilité (π_0): Kendall et Stuart ne s'y aventurent pas. Liddell (1978), le premier, propose d'y substituer l'estimateur du maximum de vraisemblance, c.-à-d. la valeur de π telle que la probabilité finale obtenue à partir des données disponibles soit la plus grande. Il s'agit naturellement de l'estimateur pondéré

$$\hat{\pi}_0 = \frac{n_{.1}p_1 - n_{.2}p_2}{n} = \frac{n_{11} + n_{12}}{n}$$

La procédure de Liddell produit ainsi une probabilité ponctuelle, en calculant :

$$p_{\text{Liddell}} = \sum_A p(n_{11}' | \hat{\pi}_0, n_{.1}) \times p(n_{12}' | \hat{\pi}_0, n_{.2}) \quad (8)$$

Poussant l'idée un peu plus loin, Laurencelle (2005) propose d'utiliser chaque valeur possible du paramètre commun π_0 en pondérant cette valeur par sa vraisemblance, celle-ci étant évidemment mesurée par :

$$\sim \pi_0^{n_{.1} + n_{.2}} (1 - \pi_0)^{n - n_{.1} - n_{.2}}.$$

Il en résulte une probabilité moyenne, fournie par :

$$p_{\text{Diff}} = \frac{(n+1)!n_{.1}!n_{.2}!}{(2n+1)!n_{.1}!n_{.2}!} \sum_{t_1, t_2 \in A} \frac{(n_{.1} + t_1 + t_2)!(2n - n_{.1} - t_1 - t_2)!}{t_1!(n_{.1} - t_1)!t_2!(n_{.2} - t_2)!} \quad (9)$$

Un exemple numérique. Les préférences de couleurs, au tableau de la figure 3, serviront d'exemple. Nous avons ici $n_{.1} = 10$ hommes, dont $n_{11} = 6$ préfèrent les couleurs mates, et $n_{.2} = 10$ femmes, avec $n_{12} = 2$ pour ces couleurs. La différence observée est ici $p = 6/10 - 2/10 = 0,40$. L'ensemble des valeurs de n_{11} et n_{12} produisant une différence égale ou plus grande (en

mode unilatéral) est :

$$A = \{ 4:0 ; 5:0 ; 5:1 ; 6:0 ; 6:1 ; \mathbf{6:2} ; 7:0 ; 7:1 ; 7:2 ; 7:3 ; 8:0 ; 8:1 ; 8:2 ; 8:3 ; 8:4 ; 9:0 ; 9:1 ; 9:2 ; 9:3 ; 9:4 ; 9:5 ; 10:0 ; 10:1 ; 10:2 ; 10:3 ; 10:4 ; 10:5 ; 10:6 \}$$

La procédure de Liddell utilise l'estimateur du maximum de vraisemblance $\hat{\pi} = (6 + 2) / (10 + 10) = 0,40$ et elle l'applique à chaque produit de Binomiales dans (8). Par exemple, pour le couple $(n_{11} : n_{12}) = (7:1)$ de l'ensemble A , le produit de Binomiales est :

$$\begin{aligned} & p(7 | 0,40, 10) \times p(1 | 0,40, 10) \\ &= {}_{10}C_7 0,40^7 0,60^3 \times {}_{10}C_1 0,40^1 0,60^9 \\ &= 0,04247 \times 0,04031 \approx 0,001712. \end{aligned}$$

La somme des produits binomiaux pour l'ensemble A donne la probabilité unilatérale $p_{\text{Liddell}} = 0,0538$.

La procédure Diff détaillée plus haut exploite le même ensemble unilatéral mais elle repose sur une moyenne pondérée du paramètre commun π_0 plutôt que sur la valeur ponctuelle $\hat{\pi}_0$. L'évaluation de (9) fournit cette fois $p_{\text{Diff}} = 0,0505$.

Approximation. En mode d'approximation normale et sous l'hypothèse de non différence entre les groupes, la différence $p_1 - p_2$ a pour espérance 0 et pour variance $\pi(1-\pi)[1/n_{.1} + 1/n_{.2}]$. Il est d'usage de substituer au paramètre inconnu π son estimateur du maximum de vraisemblance, $\hat{\pi} = (n_{.1}p_1 + n_{.2}p_2) / (n_{.1} + n_{.2}) = (n_{11} + n_{12}) / n$. L'écart-réduit classique

$$z = \frac{p_1 - p_2}{\sqrt{\hat{\pi}(1 - \hat{\pi}) \left[\frac{1}{n_{.1}} + \frac{1}{n_{.2}} \right]}} \quad (10)$$

correspond aussi au Khi-deux classique (sans correction), calculable par la formule (2)⁹. Cependant, la distribution de

⁹ Appliquée aux données de la figure 3, la formule (10) ci-

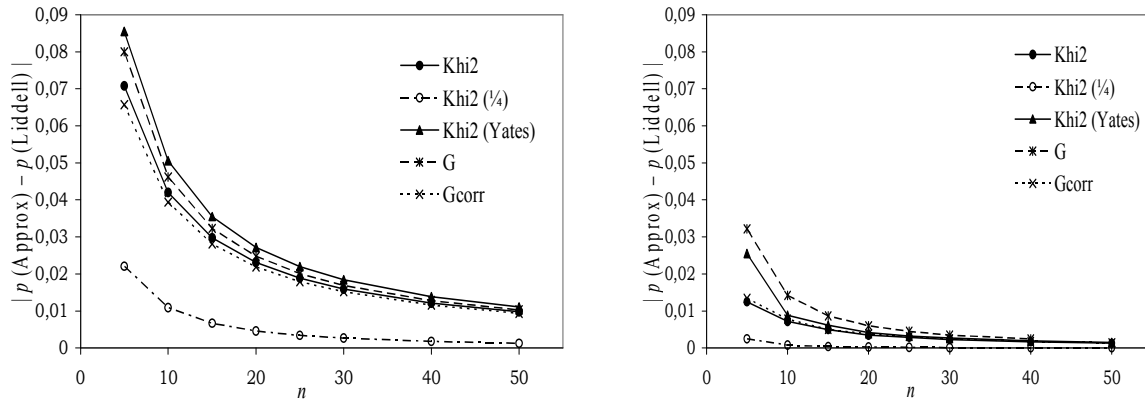


Figure 7 (gauche). Erreur moyenne (en probabilité) de 5 approximateurs du test de Liddell ($n_1 = n_2 = n$, tous p).
 Figure 7 (droite). Erreur moyenne (en probabilité) de 5 approximateurs du test de Liddell ($n_1 = n_2 = n$, $p < 0,05$)

probabilités relative à nos données de fréquences n'est pas continue, la différence $p_1 - p_2$ ayant un saut de continuité pour chaque changement d'unité dans le groupe 1 et le groupe 2. Le principe de correction de Yates (qui s'applique en principe à chaque Binomiale impliquée dans le calcul) consiste à réduire de moitié l'intervalle d'unité afin que l'aire normale ainsi définie coïncide plus précisément avec la somme des probabilités binomiales correspondantes. C'est par un tel raisonnement que Laurencelle (2005) arrive à l'approximation

$$z_{1/4} = \frac{p_1 - p_2 \pm \frac{1}{4}(1/n_{.1} + 1/n_{.2})}{\sqrt{\hat{\pi}(1 - \hat{\pi}) \left[\frac{1}{n_{.1}} + \frac{1}{n_{.2}} \right]}} \quad (11)$$

le signe appliqué pour « $\pm 1/4$ » ayant pour but et effet de réduire la valeur absolue de l'expression au numérateur. Pour nos données, cette formule s'évalue comme

$$z_{1/4} = \frac{0,6 - 0,2 - \frac{1}{4}(1/10 + 1/10)}{\sqrt{0,4 \times 0,6 \left[\frac{1}{10} + \frac{1}{10} \right]}} = 1,5975$$

valeur dont la probabilité unilatérale (sous la loi normale) est 0,0551. Notons que Pirie et Hamdan (1972) arrivent à un résultat identique pour le modèle 1, cette fois sous la forme du Khi-deux avec correction de « $-n/4$ »¹⁰.

Étude des approximations. Pour le traitement des tableaux de modèle 1 (la différence de deux proportions indépendantes), nous retenons comme référence le test de Liddell (8), plutôt que notre test Diff (9) : l'examen montre que, à travers des centaines de tableaux explorés, la corrélation des probabilités produites par ces deux tests varie de 0,9990 à 0,9999, et que les sanctions de significativité concordent à plus de 99%. Le calcul de l'une ou l'autre formule restant quelque peu encombré, l'intérêt d'une bonne approximation s'impose. Les approximateurs retenus sont :

- le Khi-deux brut (formule 2), valant 3,333 pour notre exemple ;
- le Khi-deux de Yates (formule 3), valant ici 1,875 ;
- le Khi-deux (1/4) de Pirie et Hamdan (note 10), ou notre formule $z_{1/4}$ (11), qui valent respectivement 2,552 et 1,598 ;
- le G (formule 6), ici 3,452 ;
- le G_{corr} (formule 7), soit 3,199.

Personne ne sera surpris de retrouver les deux premières formules dans les rangs des compétiteurs, puisque ce sont à toutes fins pratiques les seules que l'on retrouve partout dans les manuels d'applications statistiques. Le Khi-deux de Pirie et Hamdan (1972) concorde avec notre $z_{1/4}$ à correction réfléchie (Laurencelle 2005). Enfin, Sokal et Rohlf (1981) et d'autres recommandent généralement les tests G, en versions brute ou corrigée, pour les traitements relatifs aux modèles 0 et 1.

Les mêmes conditions de comparaison que ci-dessus ont été appliquées dans le cas présent, le résultat du test de Liddell servant cette fois de référence. Les figures 7 gauche et 7 droite affichent les différences absolues moyennes de probabilité de chaque approximateur.

En considérant tout le champ des résultats et toutes les probabilités obtenues, la figure 7 gauche fait clairement ressortir le mérite du Khi-deux(1/4), qui ressort aussi de l'analyse plus fine des probabilités liminales, à la figure 7 droite. Le test G_{corr} et, dans une moindre mesure, le Khi-deux sans correction se qualifient au second rang. La prédominance du Khi-deux(1/4) s'érode de façon importante dans le cas de groupes à tailles inégales¹¹, sans disparaître cependant, puisqu'elle se maintient jusque dans le cas extrême exploré, avec $n_2 = 5$ et $n_2 = 55$.

Les figures 8 gauche et 8 droite rapportent cette fois les cas

dessus vaut $z \approx 1,826$, et la formule (2), Khi-deux $\approx 3,333 = z^2$.

¹⁰ Dans la notation d'un tableau (a, b, c, d) , leur formule (3.3) serait $\text{Khi-deux}(1/4) = n(|ad - bc| - n/4)^2 / [(a+b)(c+d)(a+c)(b+d)]$.

¹¹ Pirie et Hamdan (1972) ont d'ailleurs élaboré leur proposition dans le contexte de groupes égaux, et la logique de la correction raisonnée de Laurencelle (2005) ne s'applique rigoureusement qu'à ces cas, l'extension à des groupes inégaux restant conjecturale.

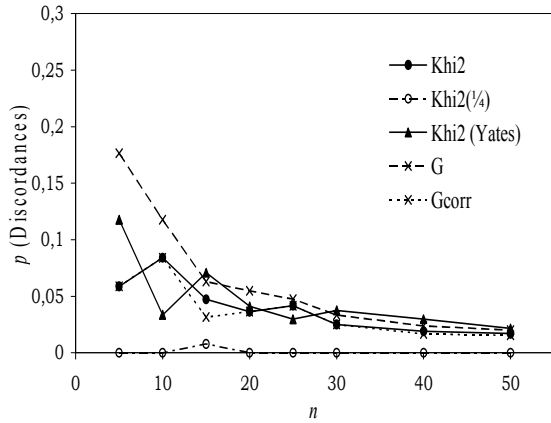


Figure 8 (gauche). Discordances de 5 approximateurs du test de Liddell ($n_1 = n_2 = n$).

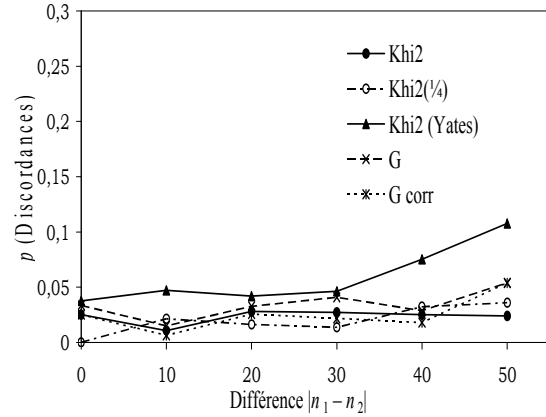


Figure 8 (droite). Discordances de 5 approximateurs du test de Liddell ($n_1 \neq n_2, n_1 + n_2 = 60$)

de discordances dans la sanction de significativité entre les différents approximateurs en lice et le test de Liddell ; rappelons que la probabilité unilatérale doit être inférieure ou égale au seuil de 0,05 pour mériter la sanction. Les tableaux pour tailles égales, représentés à la figure 8 gauche, montrent encore une fois l'avantage du Khi-deux(1/4), le compétiteur le plus consistant étant sans doute le G_{corr} . Les cas de tailles inégales rendent un verdict moins clair ; il reste cependant que le Khi-deux(1/4) apparaît souvent le meilleur, sinon l'un des meilleurs, et qu'encore une fois le G_{corr} peut prétendre à une certaine fiabilité.

La situation du modèle 1 met en comparaison deux groupes, chacun avec une probabilité (π_1, π_2) inconnue d'opter pour l'un des deux choix. Ce contexte paramétrique nous permet d'explorer une nouvelle dimension des procédures comparées, à savoir leur *puissance*. Nous avons donc généré tous les tableaux possibles découlant des tailles n_1 et n_2 et gouvernés par des valeurs paramétriques π_1 et π_2 imposées ; de plus, seuls les meilleurs compétiteurs du Liddell, soit le Khi-deux simple, le Khi-deux(1/4) et le G_{corr} , ont été retenus à cette fin. Dans ce contexte, nous avons comparé des groupes à tailles égales ou inégales, et nous avons contrasté des situations paramétriques *symétriques*, soit $\pi_1 = 1 - \pi_2$, et des situations *asymétriques*, soit $\pi_1 \neq \pi_2, \pi_1 \neq 1 - \pi_2$. Les conclusions étant toutes semblables, nous retenons à toutes fins pratiques deux cas de tailles égales et à valeurs paramétriques symétriques, tel qu'illustré dans les figures 9 gauche et 9 droite.

Les tests étant réalisés au seuil unilatéral nominal de 0,05, la puissance varie en principe de 0,05 (pour $\pi_1 - \pi_2 = 0$) jusqu'à 1 (pour $\pi_1 - \pi_2 > 0$). La parenté étroite entre le Khi-deux(1/4) et le test de Liddell réapparaît clairement à la figure 9 gauche, tout comme la belle proximité entre le G_{corr} et le Khi-deux simple, toutes les fonctions de puissance ayant tendance à se rallier dès $n_1 = n_2 = 20$.

L'avantage apparent de puissance du Khi-deux simple (et

du G_{corr}), à la figure 9 gauche, repose sur deux facteurs rédhibitoires. D'abord, dans la situation de non différence, c.-à-d. où $\pi_1 = \pi_2$, la probabilité effective de sanction significative pour ces tests est de 0,058, donc en excès du seuil imposé de 0,05. Toujours dans cette situation, l'avantage des approximateurs est obtenu en produisant des sanctions faussement significatives, en contradiction avec le test de référence. Cet avantage est, pour ainsi dire, aggravé par le fait que le calcul du test exact de Liddell opère explicitement sur des variables binomiales, donc discrètes, et que les probabilités qui en dérivent ont des sauts de continuité. Le même phénomène se produit, par exemple, dans le test binomial simple. Ainsi, pour une Binomiale avec $\pi = 1/2$ et $n = 10$, la probabilité effective de rejet au seuil de 0,05 se ramène en fait à 0,011, alors qu'avec le Khi-deux, cette probabilité monterait jusqu'à 0,055. Le chercheur doit-il respecter à tout prix le seuil de signification (α) déclaré, c'est-à-dire ne jamais le défoncer lorsque l'hypothèse nulle est posée, ou peut-il pécher un peu par optimisme, en vue de favoriser la puissance? La question, bien légitime, déborde le cadre de cet article.

Traitement statistique du modèle 0 : la corrélation entre variables dichotomiques

Théorie du traitement. Les totaux des rangées et colonnes n'étant pas prédéterminés, chacun des n éléments en jeu reste libre d'occuper n'importe laquelle des cellules a, b, c ou d du tableau, selon les probabilités p_a, p_b, p_c et p_d , d'où la probabilité d'une répartition (a, b, c, d) donnée obéit à la loi multinomiale, soit

$$p(a', b', c', d') = \binom{n}{a', b', c', d'} p_a^{a'} p_b^{b'} p_c^{c'} p_d^{d'} \quad (12)$$

ce à travers l'ensemble des configurations (a', b', c', d') pertinentes. Selon l'hypothèse de non dépendance, l'inscription de chaque élément dans une cellule résulte du produit des probabilités correspondantes en rangée et en

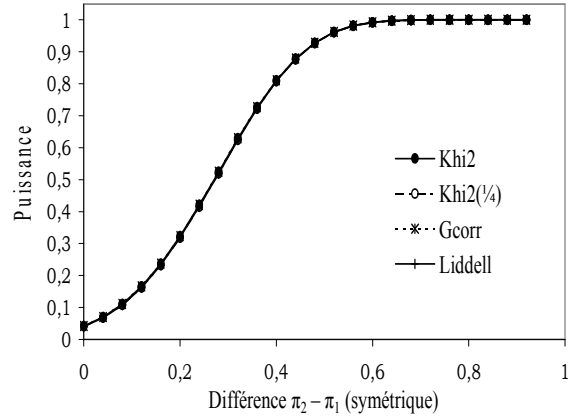
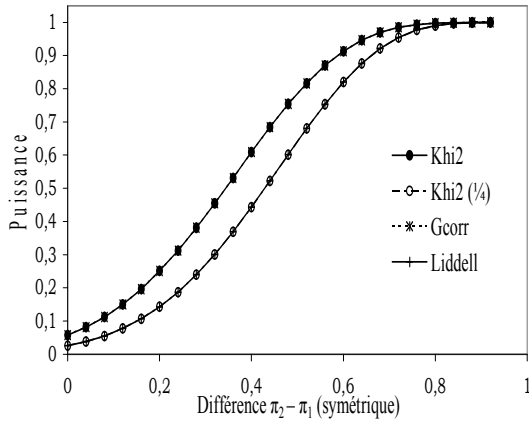


Figure 9 (gauche). Puissance de 4 tests au seuil de 0,05, pour $n_1 = 10$ et $n_2 = 10$ (différence $\pi_2 - \pi_1$ symétrique). Les courbes superposées du Liddell et du Khi2(1/4) sont au bas des deux autres.

Figure 9 (droite). Puissance de 4 tests au seuil de 0,05, pour $n_1 = 20$ et $n_2 = 20$ (différence $\pi_2 - \pi_1$ symétrique). Toutes les courbes sont superposées.

colonne. Nommons ces probabilités π_X (pour la rangée X+) et π_Y (pour la colonne Y+), alors nous avons :

$$p_a = \pi_X \pi_Y ; p_b = \pi_X(1-\pi_Y) ; p_c = (1-\pi_X)\pi_Y ; p_d = (1-\pi_X)(1-\pi_Y). \quad (13)$$

Enfin, les estimateurs du maximum de vraisemblance

$$\hat{\pi}_X = \frac{a+b}{n} ; \hat{\pi}_Y = \frac{a+c}{n} \quad (14)$$

permettent en principe de compléter l'évaluation de (12).

Quel est « l'ensemble des configurations pertinentes » mentionné ci-dessus et sur quoi ce tableau nous permet-il de conclure, au juste? À défaut d'une tradition d'interprétation ou d'un raisonnement homologué pour le cas du modèle 0 dans la documentation, nous concevons que :

1) la dépendance pouvant émerger d'un tableau à double dichotomie de type (a, b, c, d) apparaît comme une corrélation, c.-à-d. un degré de concordance marqué dans les choix manifestés en X et en Y ;

2) ce degré de concordance ou de discordance est proportionnel aux agrégats $(a+d)$ et $(b+c)$, qui reflètent, disons, les covariations positive et négative dans les choix, respectivement.

Un indice possible de corrélation¹², approprié à cette définition, serait alors

$$C = \frac{(a+d) - (b+c)}{n}, \quad (15)$$

variant entre +1 (pour une concordance de choix parfaite) à -1 (pour une discordance parfaite), en passant par zéro, indiquant l'incohérence des choix.

La définition de la statistique C, ci-dessus, réduit la

¹² Le coefficient ϕ (ou r de Galton-Pearson), $\phi = (ad-bc) / \sqrt{(a+b)(c+d)(a+c)(b+d)}$, obéit à une logique linéaire (d'intervalle) et ne convient ici qu'à titre d'approximation.

combinatoire du tableau (a, b, c, d) et son évaluation multinomiale en une combinatoire plus simple, soit un tableau binaire $[(a+d), (b+c)]$, à solution binomiale, l'ordre de variation passant de n^3 à n . En fait, sous l'hypothèse de non dépendance, la variable-agrégat $(a+d)$ a pour distribution la Binomiale :

$$a+d \sim B[\pi_X \pi_Y + (1-\pi_X)(1-\pi_Y), n] \quad (16)$$

l'ensemble des configurations pertinentes devenant simplement $x = a+d, a+d+1, \dots, n$ (pour $a+d \geq n/2$, en mode unilatéral).

Un exemple numérique. Le tableau présenté à la figure 2 servira d'exemple. Les concordances dominantes y sont $(a+d) = 28+8 = 36$, contre $(b+c) = 14$; si elle s'avère significative, cette « corrélation » suggérerait deux axes de préférences, soit un axe plutôt hédoniste (culture et plaisir), et un autre plutôt rationnel (sport et santé). L'indice C est ici égal à $(36 - 14) / 50 = 0,44$.

Les estimateurs de π_X et π_Y (14) sont, ici, de $30/50 = 0,60$ et $40/50 = 0,80$, fournissant les probabilités par cellule (13), soit $p_a = 0,60 \times 0,80 = 0,48$, $p_b = 0,12$, $p_c = 0,32$ et $p_d = 0,08$. Ainsi, pour la Binomiale (16), la probabilité de référence est $\hat{\pi} = (p_a + p_d) = 0,48 + 0,08 = 0,56$. Il s'agit donc d'évaluer les probabilités binomiales pour l'obtention de $x = 36$ concordances ou plus, sur un total de 50.

L'évaluation directe de

$$p_{\text{Corr}} = \sum_{a+d}^n \binom{n}{x} \hat{\pi}^x (1-\hat{\pi})^{n-x} \quad (17)$$

donne ici $p_C \approx 0,0149$. L'approximation normale classique du test binomial¹³, soit

¹³ Cette formule est équivalente à celle d'un test Khi-deux

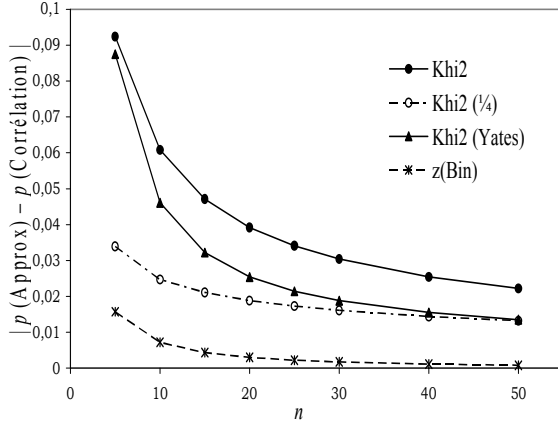


Figure 10 (gauche). Erreur moyenne (en probabilité) de 4 approximateurs du test de corrélation ($n_1 = n_2 = n$, tous p).

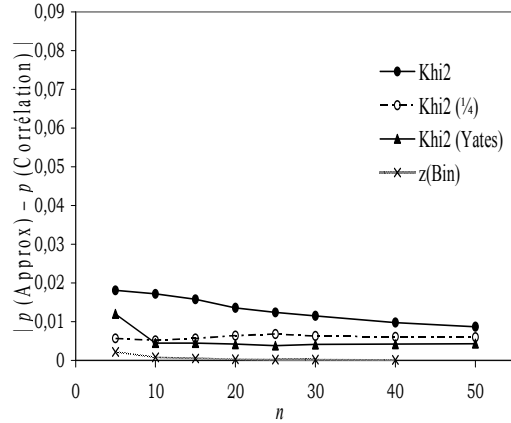


Figure 10 (droite). Erreur moyenne (en probabilité) de 4 approximateurs du test de corrélation ($n_1 = n_2 = n$, $p < 0,05$).

$$z_{\text{Bin}} = \frac{(a+d) - n\hat{\pi} \pm \frac{1}{2}}{\sqrt{n\hat{\pi}(1-\hat{\pi})}} \quad (18)$$

implique elle aussi une correction de Yates, l'application de « $\pm \frac{1}{2}$ » visant à réduire la valeur absolue de l'expression au numérateur. Elle vaut ici $z = (36 - 0,56 \times 50 - \frac{1}{2}) / \sqrt{50 \times 0,56 \times 0,44} \approx 2,137$, de probabilité unilatérale $p \approx 0,0163$. Par un test ou l'autre, il y aurait donc une connexion significative dans le tableau, une concordance marquée pour certains choix.

Approximation. Il n'existe pas d'approximations crédibles pour vérifier statistiquement la corrélation manifestée dans un tableau de fréquences 2×2 ; toutefois, d'autres mesures de corrélation existent (voir p.ex. Kendall et Stuart 1979, pp. 568-571). Parmi celles-là, on peut citer le coefficient ϕ (voir note 12), une traduction du coefficient de corrélation linéaire de Galton-Pearson pour des variables dichotomiques, et dont le test s'effectuerait selon Khi-deux = $n \times \phi^2$, restituant le Khi-deux brut (2) bien familier. Pour les données présentes, soit (28, 2, 12, 8), $\phi \approx 0,408$ et Khi-deux $\approx 8,333$, de probabilité extrême $p \approx 0,002$.

Plus sérieusement, la « corrélation tétrachorique » (Guilford 1965, Martin et Baillargeon 1989, McNemar 1969), notée r_t , estime la valeur de la corrélation ρ dans une Binormale $N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ découpée à la fois en X au quantile $\pi_x = (a+b)/n$ et en Y au quantile $\pi_y = (a+c)/n$. Kendall et Stuart (1979) en détaillent le procédé d'estimation, très complexe. Son application donne ici $r_t \approx 0,667$. La documentation fournit aussi une approximation de l'erreur-type de r_t sous l'hypothèse nulle, soit $\sigma(r_t) \approx [\pi_x \pi_y (1-\pi_x)(1-\pi_y)/n]^{1/2} / (f_x f_y)$, où, p.ex., f_x est l'ordonnée de la fonction de densité normale standard au quantile π_x . Pour nos données,

$\sigma(r_t) = [0,6 \times 0,8 \times 0,4 \times 0,2 / 50]^{1/2} / (0,3863 \times 0,2800) \approx 0,256$. Le test par approximation normale serait alors $z = r_t / \sigma(r_t) = 0,667 / 0,256 \approx 2,605$, de probabilité $p \approx 0,005$: ces estimations de probabilité n'ont fait l'objet d'aucune étude, à notre connaissance. Noter que la documentation propose aussi une assez bonne approximation de la valeur de r_t par la formule dite « cos-pi », soit

$$r_t \approx \cos \left(\pi \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \right) = \cos \left(\frac{\pi}{1 + \sqrt{ad/bc}} \right)$$

qui vaut ici 0,715. Encore une fois, cependant, rien sur la significativité d'un tel résultat.

Étude des approximations. Le concept de corrélation que nous mettons de l'avant, notamment par le ratio de concordances et discordances (15), n'a pas été considéré comme tel dans la documentation. Il reste que les auteurs – ceux qui distinguent le modèle 0 comme ceux qui l'ignorent – proposent des procédures de test : Kendall et Stuart (1979) plaident pour la valeur générale du Khi-deux simple (dans un contexte asymptotique), Siegel et Castellan (1988) préfèrent le Khi-deux de Yates ou, pourquoi pas, le Fisher, tandis que Sokal et Rohlf (1981) appliquent partout le G sauf dans le cas du modèle 2. Les approximateurs que nous retenons ici sont :

- le Khi-deux brut (formule 2), soit 8,333 pour notre exemple ;
- le Khi-deux de Yates (formule 3), ici 6,380 ;
- le Khi-deux (1/4) de Pirie et Hamdan (note 10), ici 7,324 ;
- le z(Bin) (formule 18), soit 2,137 tel qu'indiqué plus haut,

notre test Corr (17) servant de référence.

Les figures 10 gauche et 10 droite font ressortir les mérites comparés des quatre candidats ; les conditions de comparaison reconduisent celles appliquées aux modèles précédents. Bien sûr, le test binomial, z(Bin), s'en tire avec les honneurs, comme on devait l'escompter de cette fameuse

d'ajustement, dans lequel les fréquences observées $a+d$ et $b+c$ seraient respectivement contrastées avec les fréquences théoriques, $n\hat{\pi}$ et $n(1-\hat{\pi})$.

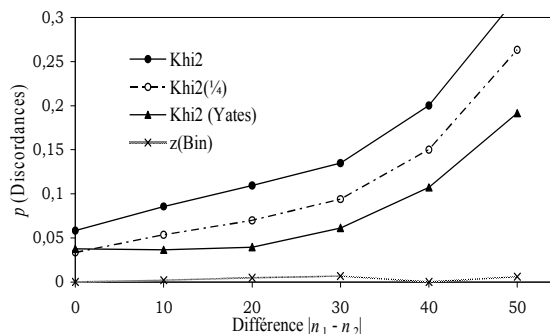
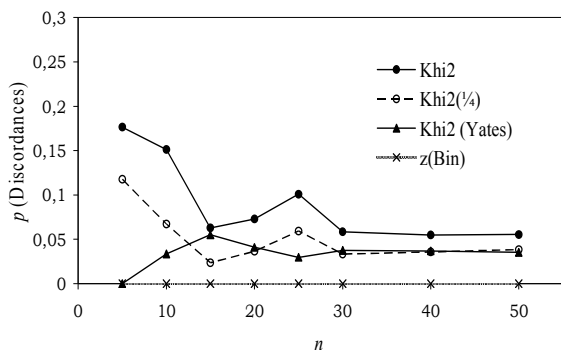


Figure 11 (gauche). Discordances de 4 approximateurs du test de corrélation ($n_1 = n_2 = n$).

Figure 11 (droite). Discordances de 4 approximateurs du test de corrélation ($n_1 \neq n_2, n_1 + n_2 = 60$).

approximation remontant à De Moivre. Il reste que les procédures basées sur le Khi-deux ne sont pas à dédaigner, en particulier le Khi-deux($1/4$), dont l'erreur moyenne dépasse à peine celles des meilleurs candidats déjà couronnés pour les modèles 2 et 1.

Les qualités relatives des approximateurs sont révélées aussi dans les figures 11 gauche et 11 droite, qui illustrent le niveau de discordances de chaque procédure par rapport à notre Corr (17).

Excepté le comportement pour ainsi dire parfait de l'approximation classique $z(\text{Bin})$, on voit, surtout à la figure 11 droite, que les autres procédures résistent mal à la présence d'une asymétrie (i.e. $n_1 \neq n_2$) dans le tableau 2×2 . N'eût été de cette déconvenue, le Khi-deux de Yates ou celui de Pirie et Hamdan à correction « $-n/4$ » auraient été dignes de considération.

Conclusion et recommandations générales

Pour autant qu'il puisse identifier le modèle probabiliste auquel ressortit son tableau de fréquences 2×2 , le statisticien appliqué a à sa disposition un répertoire suffisant de tests appropriés, dont nous croyons avoir démontré la spécificité.

Les choix totaux sont prédéterminés. Dans le tableau analysé, les totaux de rangées (p. ex. les groupes de tailles n_1 et n_2) et ceux de colonnes (p. ex. deux catégories d'options, à capacités prédéfinies) peuvent être fixés d'avance, à l'instar de notre modèle 2. C'est alors le test à probabilités hypergéométriques de Fisher (5) qui s'applique précisément, le Khi-deux de Yates (à correction de « $-n/2$ ») étant de loin sa meilleure approximation.

La comparaison de deux proportions indépendantes. Le test de Liddell (8) fut conçu spécifiquement pour les situations où les participants de deux groupes doivent opter librement pour l'une ou l'autre de deux catégories, c.-à-d. pour tester la différence entre deux proportions indépendantes ; le test Diff (9) en constitue une autre opérationnalisation. Il s'agit ici de notre modèle 1, soit, à ce que nous croyons, le paradigme d'utilisation le plus répandu pour les tableaux 2×2 . L'une et

l'autre procédure du test s'avérant un peu lourdes, il est heureux que le Khi-deux($1/4$) de Pirie et Hamdan (note 10), ou sa version $z_{1/4}$ (11) plus explicite, apparaisse comme une excellente approximation, substantiellement en avance par rapport aux procédures de test (Khi-deux simple ou de Yates, tests G simple ou corrigé) habituellement appliquées.

La corrélation entre deux variables dichotomiques. Le cas du modèle 0, aussi appelé « double dichotomie » dans la documentation (p. ex. Kendall et Stuart 1979), a été traité ici en mode exploratoire. À notre su, le traitement spécifique de ce modèle n'a jamais été publié, sauf à titre descriptif et en référence à la corrélation linéaire (par le coefficient ϕ ou la corrélation tétrachorique r_t). Pour déterminer s'il y a corrélation significative des choix dans un tableau 2×2 , nous avons proposé un indice de concordance, l'indice C (15), ainsi qu'un test axé sur le ratio des concordances / discordances présentes dans le tableau, soit le test Corr (17), test dont l'évaluation binomiale exacte est très facilement approximée par l'écart-réduit normal classique, $z(\text{Bin})$ (18). L'avenir dira si cette opérationnalisation de l'information dans un tableau 2×2 va survivre.

Quant à la valeur des approximations elles-mêmes, il s'est avéré que le Khi-deux brut (2) n'a pas de lieu de convenance dans le traitement des tableaux 2×2 , non plus que les tests du rapport de vraisemblance, G (6) et G_{corr} (7). Seuls les Khi-deux corrigés, soit selon Yates avec correction de « $-n/2$ » (3), soit selon Pirie et Hamdan et Laurencelle avec correction de « $-n/4$ », survivent à l'examen. Ce n'est pas dire que les tests sacrifiés n'ont pas d'application. Le traitement des tableaux de fréquences $2 \times k$ ou $r \times k$ exige ces tests, que Sokal et Rohlf (1981) proposent à juste titre, et les propriétés de décomposition additive particulièrement avantageuses du test G (voir aussi Black et Laurencelle 1987) le recommandent à notre attention. Rappelons enfin que le soi-disant « test exact de Fisher », en fait un test à probabilités hypergéométriques, n'est « exact » que pour le traitement des tableaux du modèle 2, un cas rare, et qu'il peut au mieux servir d'approximation, une approximation douteuse, dans les autres cas.

References

- Black, P., Laurencelle, L. (1987). Le test G pour les tableaux de fréquences et sa décomposition additive. *Lettres Statistiques*, 8, 99-114.
- Cochran, W. C. (1952). The χ^2 test of goodness of fit. *Annals of mathematical statistics*, 23, 315-345.
- Conover, W. J. (1974). Some reasons for not using the Yates continuity correction on 2x2 contingency tables. *Journal of the American Statistical Association*, 69, 374-376.
- Grizzle, J. E. (1967). Continuity correction in the X^2 -test for 2 x 2 tables. *American Statistician*, 21, 28-32.
- Guilford, J. P. (1965). *Fundamental statistics in psychology and education* (4^e édition). New York : McGraw-Hill.
- Haber, M. (1980). A comparison of some continuity corrections for the Chi-square test on 2x2 tables. *Journal of the American Statistical Association*, 75, 510-515.
- Howell, D. C. (1998). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck.
- Kendall, M. G., Stuart, A. (1979). *The advanced theory of statistics. Vol. 2 : Inference and relationship* (4^e édition). New York : Macmillan.
- Laurencelle, L. (2005). Le test binomial exact de la différence entre deux proportions et ses approximations. *Lettres Statistiques*, 12, 1-13.
- Liddell, D. (1978). Practical tests of 2x2 contingency tables. *The Statistician*, 25, 295-304.
- McNemar, Q. (1969). *Psychological statistics* (4^e édition). New York : Wiley.
- Martin, L., Baillargeon, G. (1989). *Statistique appliquée à la psychologie* (2^e édition). Trois-Rivières : SMG.
- Pirie, W. R., Hamdan, M. A. (1972). Some revised continuity corrections for discrete distributions. *Biometrics*, 28, 693-701.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York : McGraw-Hill.
- Siegel, S., Castellan, J. (1988). *Nonparametric statistics for the behavioral sciences*. New York : McGraw-Hill.
- Sokal, R. R., Rohlf, F. J. (1981). *Biometry* (2^e édition). New York : Freeman.
- Williams, D. A. (1976). Improved likelihood ratio tests for complete contingency tables. *Biometrika*, 63, 33-37.
- Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement. Journal of the Royal Statistical Society*, 1, 217-235.

Manuscript received March 27th 2007

Manuscript accepted April 2nd 2007